

# Visibilidad y uso de colecciones digitalizadas: propuesta de un indicador de visibilidad relativa

*Visibility and use of digital collections: proposing an indicator of relative visibility*

Yanet Elisa RICO PÉREZ, Michela MONTESI

Facultad de Ciencias de la Documentación de la Universidad Complutense de Madrid, Calle Santísima Trinidad 37, Madrid  
{yerico@estumail|mmontesi@pdi}.ucm.es

## Resumen

Se examinan los conceptos de visibilidad y uso para colecciones digitales y se propone una métrica que relaciona ambos conceptos. Utilizando el análisis de los registros de uso del sitio web, se propone un índice de visibilidad relativa, que consiste en el cociente de los usuarios nuevos respecto a los que retornan. Si los nuevos usuarios están muy por encima de los veteranos, implicaría que existen problemas de usabilidad, accesibilidad, contenido, etc., pues los usuarios no regresan. Si, por el contrario, los usuarios veteranos están muy por encima de los nuevos, el problema sería de visibilidad, pues muy pocos usuarios llegan por primera vez. La métrica se validó con los registros de uso de la colección digital del fondo antiguo de la Biblioteca Universidad Complutense de Madrid durante el periodo de enero a junio de 2011.

**Palabras clave:** Visibilidad web. Uso. Colecciones digitales. Repositorios. Métrica. Análisis web. Hathitrust. Biblioteca digital. Google Books.

## Abstract

The concepts of visibility and use for digital collections are discussed, and a metric is proposed that relates both of them. Using a web-analytics tool for the data collected from a website, a relative visibility index is suggested, which results from dividing the new users by those returning to the web site. If the number of new users is well above that of the previous ones, that could mean that there are problems of site usability, accessibility, content, etc., because users are not returning. If, on the contrary, the number of previous users is well above that of the new ones, there could be a problem of visibility, as fewer users are arriving to the web site. The metric was developed in a case study involving the log records of the antique book digital collection of the Universidad Complutense of Madrid, from January to June of 2011.

**Keywords:** Web visibility. Use. Digital collections. Repositories. Metrics. Web analytics. Hathitrust Digital Library. Digital libraries. Google Books.

## 1. Introducción

Examinar la visibilidad de un sitio web o de una colección digital disponible en la web no solo es un asunto de gran relevancia y actualidad para quienes pretenden que su sitio o colección digital llegue al mayor número posible de usuarios a lo largo y ancho del mundo, sino que es un tema en torno al cual se han generado diversos malentendidos. Solo por poner un ejemplo mencionemos el de pensar que la visibilidad de un sitio se hace equivalente a su popularidad.

Ahora bien, tanto del lado de los sitios web comerciales como de los académicos, la investigación sobre la visibilidad, el uso, el posicionamiento web y otros temas relacionados con la disciplina conocida como *análisis web*, ha sido objeto de investigación y producción de un número creciente de artículos desde los años 90s, cuando surgió esta disciplina, a la que la Digital Analytics Association (DAA) (s. a.) definió como “rastreo, colección, medida, reporte y análisis de datos cuantitativos de internet con el propósito

de optimizar los sitios web y las iniciativas de marketing web”. En este sentido una mirada al artículo de Penniman (2008), quien revisa la historia del nacimiento y evolución del análisis de datos log aplicado a sistemas de recuperación de la información; y, otra, al artículo de Markey (2007), quien revisa 25 años de investigación académica en el área, nos da una idea del tamaño, la importancia y la actualidad de estos temas. Así se comprende que las instituciones académicas inviertan grandes recursos en hacer que los sitios web de sus bibliotecas, sus colecciones digitales, sus publicaciones electrónicas, o sus archivos de *e-prints* no solo sean visibles, sino también usables, accesibles y, en algunos casos, que estén bien posicionados en un *ranking* global de sitios académicos.

Sentado esto, según el investigador sudafricano Melius Weideman, el concepto de visibilidad suele tener por lo menos dos acepciones. Desde un punto de vista técnico, se entiende que la visibilidad es el grado de facilidad con el cual el crawler de un motor de búsqueda puede encontrar una página Web y, una vez encontrada, se

entiende también como el grado de éxito con el cual el mismo *crawler* indexa la página (Weideman, 2009, p.14). Según algunos investigadores Españoles (Codina y Marcos 2005, p. 85), la visibilidad es un término para referirse a una medida que indica el número de enlaces que un sitio web recibe de otros sitios y la calidad de los mismos. Se desprende de estas acepciones técnicas que la participación del usuario en la visibilidad de un sitio web sería muy poca; y, según Weideman, ninguna (Weideman, 2009, p. 40).

De otro lado, el concepto de visibilidad se emplea también en un sentido más general para referirse a la visibilidad de los autores, de la producción científica, y, de las colecciones digitales, especialmente los repositorios institucionales. En este sentido, Pavão et al. (2012, p. 52), consideran la visibilidad como “el grado de exposición y la evidencia de un investigador frente a la comunidad científica”, y añaden que el papel de los repositorios es precisamente el de aumentar la visibilidad de la institución y de sus investigadores. Así, la posibilidad de obtener una mayor visibilidad sería el aliciente de los investigadores para publicar sus trabajos en los repositorios y en acceso abierto, y consecuentemente de las instituciones para invertir ingentes recursos en estos sistemas de información (Melero et al., 2009, p.20).

El aumento de la visibilidad y del número de citas representa el principal estímulo para el desarrollo de los repositorios. La visibilidad puede verse entonces como una implicación necesaria del propio concepto de repositorio que convierte a la institución en un proveedor de contenidos y la hace automáticamente visible en la web (Melero, 2008, p. 2). Sin embargo, los contenidos disponibles desde un repositorio difícilmente podrían someterse a mediciones de visibilidad según se entiende en su acepción más técnica, como vimos anteriormente.

Finalmente autores como Guallar y Abadal (2009, p. 259) incluyen la visibilidad entre los indicadores para tener en cuenta en la evaluación de hemerotecas de prensa digital y entienden la visibilidad como la “Presencia en internet de los archivos de prensa a partir del número de enlaces que apuntan a cada página en concreto.” Sin embargo, no explican si es viable aplicar este principio a conjuntos documentales de gran tamaño como pueden ser los repositorios y no tienen en cuenta que la literatura científica presenta patrones de enlace que van más allá de los hipervínculos siendo el más conocido la citación.

En el presente artículo pretendemos investigar la posibilidad de definir el concepto de visibilidad

de una colección digital teniendo en cuenta las dos acepciones de este concepto propuestas por Weideman, aunque con una diferencia. Concretamente, proponemos una medición de la visibilidad de una colección digital con base en sus patrones de uso, concepto que implica, a diferencia de Weideman, un elemento humano. Para esto apelaremos a un caso particular, a saber, un conjunto de datos de uso de la colección digitalizada del fondo antiguo de la Biblioteca de la Universidad Complutense de Madrid.

En nuestra investigación tuvimos en cuenta uno de los dos métodos mas comunes de rastreo para la recolección de datos para el análisis web (Kaushik, 2007, p. 100), a saber, la técnica conocida como etiquetado de pagina (*page tagging via JavaScript*) que es empleada, entre otros, por el software Google Analytics de la compañía Google. Esta herramienta tiene ventajas y desventajas si se la compara con otra gran técnica de rastreo para la recolección de datos, la llamada técnica de datos log (*web log files*). La técnica de rastreo *page tagging* tiene, entre otras, la ventaja de estar disponible para compañías que no tienen acceso a sus propios servidores web, como es el caso del repositorio de la Universidad Complutense que examinaremos, el cual está hospedado en el servidor de una universidad norteamericana.

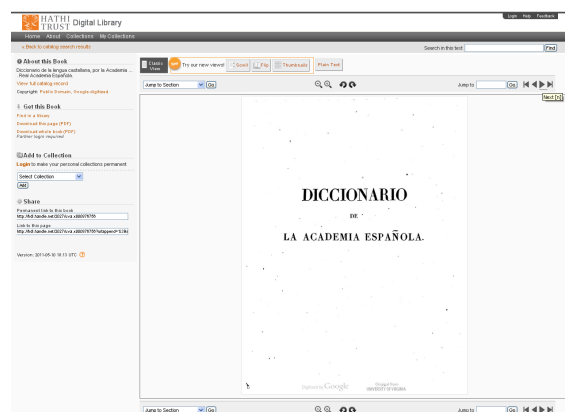


Figura 1. Interface de Hathitrust

La mencionada colección digital se realizó dentro de un proyecto de colaboración con *Google Books* y actualmente se encuentra hospedada en el archivo de Hathitrust, un consorcio de instituciones de investigación y bibliotecas principalmente americanas cuyo objetivo es la preservación de las respectivas colecciones digitales. El proyecto, realizado con *Google Books* y que comenzó en el 2006, se sitúa dentro una estrategia general de la Biblioteca de la Universidad Complutense de Madrid para diseminar el conocimiento y la investigación, construyendo

colecciones a través de la digitalización (Magán et al., 2011, p. 52).

Aproximadamente 50 bibliotecas forman parte del consorcio HathiTrust cuyo objetivo principal es servir como un repositorio que garantice la preservación a largo término de los materiales que hospeda. La biblioteca de la Universidad Complutense de Madrid aporta el 1% del contenido total del repositorio.

Desde hace más de diez años, se han iniciado varios proyectos de definición de estadísticas e indicadores de rendimiento de productos y servicios electrónicos en procesos de evaluación bibliotecaria. El proyecto Counter (*Counting Online Usage of Networked Electronic Resources*), el más conocido, puesto en marcha en 2002, ha propuesto varias guías de uso (cada una mejorando a la anterior) en las cuales establece los datos que se han de medir, las definiciones, el contenido y el formato de los informes que se han de entregar. Muchas editoriales científicas se ajustan a los criterios establecidos por Counter en los informes que entregan periódicamente a las bibliotecas. Sin embargo, las colecciones digitales creadas por las propias bibliotecas no cuentan aún con una batería de parámetros normalizados —o por lo menos consensuados— para medir el uso que se hace de ellas.

En cuanto a los repositorios, varios proyectos de investigación están trabajando para la implementación de mecanismos de medición de contenidos y tráfico. El objetivo de PIRUS (*Publisher and Institutional Repository Usage Statistics*), por ejemplo, es precisamente el de elaborar un estándar compatible con Counter que permita obtener informes de uso a nivel de artículo individual de repositorios (Brody et al. 2009). La posibilidad de obtener datos de uso a nivel del documento individual es sin duda una de las principales ventajas que ofrecen los repositorios frente a las mediciones tradicionales de impacto de las publicaciones científicas, y la multiplicación de datos susceptibles de análisis estadístico sobre publicaciones científicas ha sido un importante aliciente para la promoción del acceso abierto.

## 2. La visibilidad de la colección digital de la Universidad Complutense

La medición de la visibilidad que proponemos en este artículo se basa en datos de uso de tipo de visitante (visitante nuevo ó primera visita, y visitante que regresa) recogidos para los meses de enero y junio del 2011, para la colección digital de la BUC hospedada en HathiTrust. HathiTrust monitoriza los datos de acceso al

repositorio completo utilizando una herramienta gratuita, *Google Analytics* (GA). La aplicación *Google Analytics* genera estadísticas detalladas del tráfico a un sitio web. Se trata de un servicio de métricas web muy utilizado actualmente, pues hace uso de esta herramienta el 57% de los 10.000 sitios web más populares.

Sin embargo, aunque herramientas de análisis de sitios web como *AwStats* (que usa archivos *log*) o *Google Analytics* (que usa *page tagging*) representan una solución común para recoger datos de uso, la literatura especializada ha destacado que no se pueden aplicar de forma fiable a los repositorios (Carr et al., 2008, p. 1). En *Google Analytics* (y cualquier otro software que use la técnica de etiquetado de página) las páginas que no correspondan a archivos HTML no producen visitas cuando el usuario accede a ellas. Aunque Lee (2011) lista 11 posibles aplicaciones de GA para la evaluación de bibliotecas digitales, precisa que GA es incapaz de registrar acciones como, por ejemplo, descargar un archivo, o presionar el botón *play* de un video, o interactuar con elementos de *Ajax* o *Flash*. Cabe mencionar también que los software que usan log files adolecen entre otras limitaciones de que las páginas que quedan almacenadas en la caché web no son contadas por el servidor y, por tanto, estas páginas no contadas sesgan seriamente las métricas. Sobre las ventajas y desventajas de estas dos técnicas la obra de Kaushik (2007) es muy detallada.

Por lo que se refiere a HathiTrust es posible consultar el texto completo de una obra desde la página de su registro, y opcionalmente desde ese mismo registro descargar el texto completo en PDF de la misma. Con lo cual, *Google Analytics* registra los accesos al texto completo.

Las métricas y conceptos utilizados en *Google Analytics* se encuentran explicados en el propio sitio web de la herramienta, aunque para este trabajo hemos recurrido a un popular manual de uso, el de Ledford et al. (2010). Según este manual, las visitas a un sitio web que cuenta *Google Analytics* se realizan, en general, por “personas”, aunque lo que realmente GA cuenta en primer lugar son direcciones IP y para análisis posterior solo tiene en cuenta los cookies que asigna al navegador con el cual el usuario accedió al sitio. No obstante, hay muchas visitas a un sitio web que realmente no provienen de personas sino de “no-personas”, como las aplicaciones (arañas, crawlers, robots) a los que se le asigna la tarea de revisar un sitio web por alguna razón. A diferencia de la técnica de archivos log, la técnica de etiquetado de página empleada por *Google Analytics* no reconoce estas visitas no humanas que pueden ser de

utilidad a la hora de hacer optimización de motores de búsqueda, elemento fundamental para mejorar la visibilidad de un sitio web.

Las métricas más importantes usadas por GA, incluyen entre otros (Ledford et al. (2010.pp. 83):

- *Visitas*: Una visita es una interacción de un individuo con un sitio web, la cual consiste de una o más solicitudes de unidades de contenido (página vista).
- *Visitante único*: Es el número de personas individuales dentro de un marco de tiempo registrado con una actividad consistente en una o más visitas a un sitio. Cada individuo es contado teniendo en cuenta el IP de su ordenador en la métrica de visitante único para el periodo reportado.
- *Visitante nuevo*: Es el número de visitantes únicos con una actividad que incluye una primera visita a un sitio durante el periodo de tiempo reportado.
- *Visitante que retorna*: El número de visitantes únicos con una actividad consistente de una visita a un sitio durante un periodo de tiempo reportado, que también visitaron el sitio en periodos anteriores.

Cabe anotar que estas métricas de GA tienen limitaciones. A diferencia del análisis log de transacciones el cual usa las direcciones IP para distinguir usuarios, el Google Analytics instala un cookie en el navegador del visitante en su primera visita para comenzar a rastrear a este visitante único particular. Este cookie persistente se actualiza a sí mismo con cada visita que retorna (suponiendo que el usuario utilice el mismo navegador) y con cada página vista, hasta por dos años. Sin embargo, como los cookies pueden ser fácilmente borrados, entonces un usuario podría llegar a ser contado incorrectamente como otro visitante único con un cookie diferente en la próxima vez que haga una visita. Otros escenarios que dan como resultado malos cálculos ocurren cuando los usuarios utilizan dispositivos múltiples o navegadores múltiples, tales como un ordenador para trabajar y otro desde casa, o cuando múltiples usuarios acceden al mismo ordenador en colegios o bibliotecas. Así en casos como este en lugar de calcular un visitante único lo que se está calculando es un navegador único. En el caso de la colección digital Complutense lo más probable es que los usuarios accedan desde ordenadores distintos a los de una biblioteca, pues se trata de una colección de acceso gratuito.

### 3. Los Datos

Fue imposible obtener de forma individualizada todos los datos de uso relativos a la colección UCM, pues las estadísticas recopiladas se referían a todas las colecciones hospedadas en HathiTrust. Las mediciones que pudimos recoger fueron: tipo de visitante (visitante nuevo, visitante que retorna), fuentes de tráfico, idioma, ciudad, país y palabra clave.

En este artículo nos centraremos en las métricas de visitantes nuevos y visitantes que retornan, observando que en los conjuntos de datos que habíamos recogido para enero y junio de 2011 ambas categorías de visitas tendían a aumentar en los seis meses. Se trata de una evolución normal dado que el tamaño de la colección también aumentó y, de hecho, fue en el mes de mayo de 2011 cuando se concluyó la ingesta de textos digitalizados complutenses a HathiTrust (Magán et al., 2011, p. 62). Un aspecto que nos llamó la atención fue que la proporción de visitantes que retornaban con respecto a los nuevos fue aumentando de enero a junio hasta alcanzar un nivel de equilibrio. Interpretamos los visitantes nuevos como aquellos que *ven* la colección digital, llegando a ella a través de motores de búsqueda o enlaces desde otros sitios web. Por otro lado, entendimos que aquellos visitantes que retornan están utilizando la colección, pues ya la conocen y regresan para realizar alguna tarea con los documentos que contiene.

En el Gráfico 1 se muestra el comportamiento de los visitantes nuevos para los primeros 30 días de los meses de enero y junio. Análogamente, el Gráfico 2 muestra una superioridad de lo ocurrido en junio respecto de enero para los visitantes que retornan.

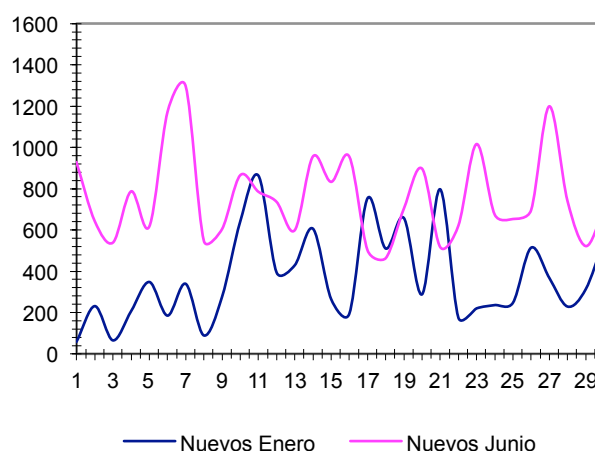


Gráfico 1. Evolución de los visitantes nuevos en enero y junio de 2011



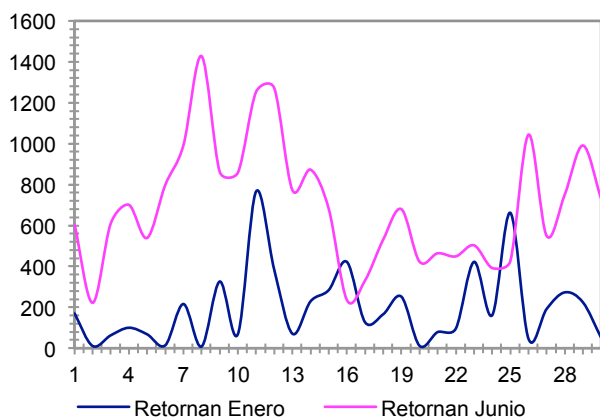


Gráfico 2. Evolución de los visitantes que retornan en enero y junio de 2011

El Gráfico 3 compara los totales de los visitantes nuevos vs. retornan para los meses de enero y junio.

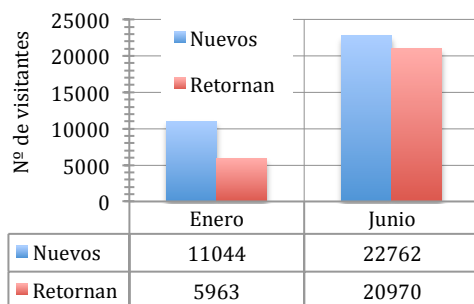


Gráfico 3. Comparación de visitantes totales nuevos y veteranos en enero y junio de 2011

Es posible observar una tendencia de los valores de enero y junio hacia la equivalencia, que nos sugiere una especie de situación óptima para la colección digital que atrae aproximadamente igual número de visitantes nuevos que de visitantes que retornan.

Esta tendencia hacia la equivalencia puede conceptualizarse en una fórmula a la que denominamos Índice de Visibilidad Relativa (IVR). Esta medida tiene la característica de estar basada solo en dos tipos de datos de tráfico web: el número de visitantes nuevos y el número de visitantes que retornan.

Atendiendo a dos situaciones hipotéticas extremas presentadas en los gráficos 4 y 5, se puede concluir, por un lado, que, si el número de visitantes nuevos en un determinado período está muy por encima respecto al número de visitantes que retornan (Gráfico 4), entonces la colección digital, en ese período, pese a ser muy visible, ha experimentado algún tipo de problema, pues los usuarios que ya la conocen prácti-

camente no regresan a ella, de donde se puede concluir que el sitio tiene problemas de usabilidad, de accesibilidad, de pobreza de contenido, u otros dependiendo de las circunstancias.

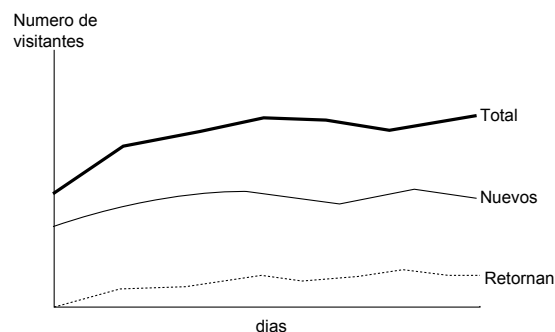


Gráfico 4. Hipotético caso extremo de niveles de uso por tipo de visitante nuevo en un periodo

De otro lado, si el número de usuarios nuevos en un determinado período está muy por debajo del número de los que retornan (Gráfico 5), entonces la colección, en ese período, pese a ser muy atractiva para quienes ya la conocen (pues regresan en masa), ha experimentado un problema de visibilidad, pues hay muy pocos usuarios que llegan por primera vez.

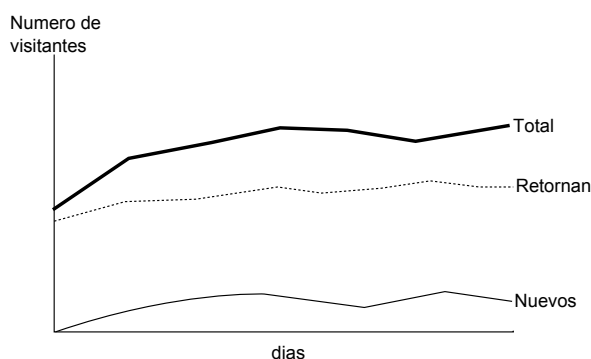


Gráfico 5. Hipotético caso extremo de niveles de uso por tipo de visitante que retorna en un periodo.

Ninguna de estas dos situaciones parece buena para una colección digital. Consideramos que lo óptimo es que el número de visitantes nuevos y que retornan, en el largo tiempo, sea aproximadamente el mismo: Así pues, cuantificaremos el Índice de Visibilidad Relativa (IVR) de acuerdo a la siguiente fórmula:

$$IVR = (\# \text{Total nuevos}) / (\# \text{Total retornan})$$

La tendencia en una colección documental digital de dominio público, es que este índice se mantenga próximo a 1

$$(0 < IVR < 1 \text{ ó } IVR > 1) \text{ en el largo tiempo}$$

Más aún, si entendemos a los usuarios nuevos como una medida de visibilidad del sitio y a los que retornan como una medida de uso del mismo sitio, el IVR puede darnos una idea de la visibilidad del sitio a través de datos de uso relativos al mismo.

Volviendo a los datos del Gráfico 3, podemos cualificar el índice de visibilidad relativa (IVR) de la siguiente manera: Si  $IVR \geq 1$  diremos que la visibilidad de la colección digital hospedada en el sitio es positiva, es decir, que el sitio atrajo a más visitantes nuevos respecto a los que retornan. En el caso que  $IVR < 1$ , diremos que la visibilidad es negativa; es decir, el sitio atrajo menos visitantes nuevos respecto de los que retornan.

De esta forma, los índices IVR para enero y junio son:

- Enero:  $IVR = 11044/5963 = 1.85 > 1$
- Junio:  $IVR = 22762/20970 = 1.08 > 1$

de donde se concluye que aún cuando la visibilidad en ambos meses es positiva, en el mes de junio dicho índice está más próximo a 1 (que es la estabilidad esperada en el largo tiempo).

Nos atrevemos a afirmar (aunque no disponemos de todos los datos), que el IVR para la Colección Digital UCM hospedada en HathiTrust a lo largo de los primeros seis meses corridos de 2011 tiende a 1, lo cual, según nuestra definición, no solo es un buen indicador de visibilidad, sino de que el repositorio es igual de atractivo para los usuarios que ya lo conocen. Aún queda por demostrar este hecho en futuros trabajos y compararlo con otros IVR para otros repositorios.

## 4. Conclusiones

En esta investigación hemos analizado los conceptos de visibilidad y uso de colecciones digitales, examinando estos conceptos mediante datos de uso de la colección digitalizada Fondo Antiguo de la UCM hospedada en el sitio HathiTrust. Como resultado, propusimos una medida de la visibilidad de una colección digital con base en datos específicos de uso. Esta medida tiene en cuenta elementos humanos en la definición del concepto de visibilidad, en alternativa a la concepción técnica de la misma que la concibe, o bien en términos de números de enlaces, o bien en términos de hallazgo e indexación de una página por un *crawler*.

Una de las limitaciones principales de este trabajo ha sido la restricción a datos de uso correspondientes a solo dos meses del año 2011.

La recolección de los datos dependía de que se nos fuera permitido, en cada oportunidad necesaria, el acceso como administradores a la sesión de *Google Analytics* propiedad de la dirección de la biblioteca de la UCM; y la minería, al haber sido realizada sobre datos relativos al repositorio en su globalidad, exportados a hojas de Excel, requirió de un largo procesamiento de varios cientos de hojas de cálculo sin más recurso que a las funciones básicas de conteo y suma de la aplicación Excel.

Para comprobar la utilidad de este concepto de visibilidad relativo al uso, se requiere de muchos más datos, no solamente de la colección que hemos investigado, sino de otras colecciones hospedadas en otros repositorios. No obstante, esta es una primera propuesta que pretende motivar la investigación en torno a la definición de estadísticas e indicadores de uso de recursos electrónicos basada en el análisis de ficheros transaccionales; y, en este sentido, consideramos que esta nueva métrica puede ser una aportación valiosa en la medida en que sea objeto y herramienta de futuros trabajos de investigación.

## 5. Referencias

- Borrego, A.; Urbano, C. (2005). Estadísticas e Indicadores de Rendimiento de Colecciones y Servicios Bibliotecarios de Carácter Electrónico. // *El Profesional de la Información*. 14:1, 30-38.
- Bernal, I.; Pemau-Alonso, J. (2010) Estadísticas para repositorios: sistema métrico de datos en Digital. CSIC. // *El Profesional de la Información*, 19:5, 534-543.
- Brody, T.; Gedy, R.; MacIntyre, R.; Needham, P.; Pentz, E.; Rumsey, S.; Shepherd, P. (2009). Developing a global standard to enable the recording, reporting and consolidation of online usage statistics for individual journal articles hosted by institutional repositories, publishers and other entities. Final port. [http://www.jisc.ac.uk/media/documents/programme/s/pals3/pirus\\_finalreport.pdf](http://www.jisc.ac.uk/media/documents/programme/s/pals3/pirus_finalreport.pdf) (24/05/2012).
- Carr, L.; Brody, T.; Swan, A. (2008). Repository statistics: what do we want to know?. // *Third International Conference on Open Repositories*. 1:4 (April 2008). Southampton, United Kingdom.
- Codina, L.; Marcos, C. (2005). Posicionamiento Web. Conceptos y Herramientas. // *El Profesional de la Información*. 14:2 (2005) 84 - 99.
- Digital Analytics Association (s. a.). Wakefield: DAA. <http://www.digitalanalyticsassociation.org/>.
- Guallar, Javier; Abadal, Ernest (2009). Evaluación de hemerotecas de prensa digital: indicadores y ejemplos de buenas prácticas. // *El profesional de la información*. 18: 3 (mayo-junio2009) 255-269.
- Kaushik, A. (2007) *Web Analytics: An Hour a Day*. Indianapolis: Wiley, 2007.
- Ledford, Jerri; Teixeira, Joe; Mary Tyler (2010). *Google Analytics*. Third edition. Indiana: Wiley publishing Inc., 2010.
- Lee, H. J. (2011). *Google Analytics for Digital Library Evaluation*. Master Thesis. Oslo: International Master of Digital Library Learning, University of Os-

- lo. [https://oda.hio.no/jspui/bitstream/10642/987/2/Lee\\_Hsin\\_Ju.pdf](https://oda.hio.no/jspui/bitstream/10642/987/2/Lee_Hsin_Ju.pdf) (11/06/2012).
- Magán, J.A.; Palafox, M.; Tardón, E.; Sanz, A. (2011). Mass Digitization at the Complutense University Library: Access to and Preservation of its Cultural Heritage. // *Liber Quarterly*. 21:1 (2011) 48-68.
- Markey, K. (2007). Twenty-five years of end-user searching, part 1: Research findings. // *Journal of the American Society for Information Science and Technology*. 58, 1071-1081. doi:10.1002/asi.20462.
- Melero, R. (2008). El paisaje de los repositorios institucionales open access en España. // *BiD*. 20 (junio. 2008) 1-20.
- Melero, R.; Abadal, E.; Abad, F.; Rodríguez-Gairín, J. M. (2009). Situación de los repositorios institucionales en España: informe 2009. [http://handle.digital.csic.es/bitstream/10261/11354/1/Informe2009-Repositorios\\_0.pdf](http://handle.digital.csic.es/bitstream/10261/11354/1/Informe2009-Repositorios_0.pdf) (13/04/2012).
- Ortega, J. L.; Aguillo, I. (2009) Minería del uso de webs. // *El profesional de la información*. 18: 1 (enero-febrero 2009) 1-7.
- Pavão, C. G.; da Costa, J.; Horowitz, Z. Ferreira, M.; Caregnato, S. (2012). Contribución del acceso abierto a la visibilidad de la literatura científica en una institución de educación superior. // *e-colabora*. 2012, vol. 2:3, pp. 48-65.
- Penniman, W. D. (2008) Historic Perspective of Log Analysis. // *Handbook of Research on Web Log Analysis* / B. J. Jansen, A. Spink, and I. Taksa, Eds. Hershey, PA: IGI, 2008. 18-38.
- Weideman, Melius. (2009). *Website Visibility: The Theory and Practice of Improving Rankings*. Oxford: Chandos, 2009.

---

Enviado: 2012-09-03. Segunda versión: 2013-08-24.

Aceptado: 2013-08-25.

---

